View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Edinburgh Research Explorer

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

# Non-backtracking walk centrality for directed networks

**Citation for published version:**
Arrigo, F, Grindrod, P, Higham, DJ & Noferini, V 2018, 'Non-backtracking walk centrality for directed networks' Journal for Complex Networks, vol. 6, no. 1, pp. 54-78. DOI: 10.1093/comnet/cnx025

**Digital Object Identifier (DOI):**
10.1093/comnet/cnx025

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
Journal for Complex Networks

**General rights**
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 09. May. 2019

*IMA Journal of Complex Networks* (2017) Page 1 of 24
doi:10.1093/comnet/xxx000

# Nonbacktracking Walk Centrality for Directed Networks

FRANCESCA ARRIGO*
*Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK, G1 1XH.*
*Corresponding author: francesca.arrigo@strath.ac.uk

PETER GRINDROD
*Mathematical Institute, University of Oxford, Oxford, UK, OX2 6GG.*

DESMOND J. HIGHAM
*Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK, G1 1XH.*

AND

VANNI NOFERINI
*Department of Mathematical Sciences, University of Essex, Colchester, UK, CO4 3SQ.*

[Received on 2 June 2017]

The theory of zeta functions provides an expression for the generating function of nonbacktracking walk counts on a directed network. We show how this expression can be used to produce a centrality measure that eliminates backtracking walks at no cost. We also show that the radius of convergence of the generating function is related to the spectrum of a three-by-three block matrix involving the original adjacency matrix. This gives a means to choose appropriate values of the attenuation parameter. We find that three important additional benefits arise when we use this technique to eliminate traversals around the network on the grounds that they are unlikely to be of relevance. First, we obtain a larger range of choices for the attenuation parameter. Second, a natural approach for determining a suitable parameter range is invariant under the removal of certain types of nodes, so we can gain computational efficiencies through reducing the dimension of the resulting eigenvalue problem. Third, the dimension of the linear system defining the centrality measures may be reduced in the same manner. We show that the new centrality measure may be interpreted as standard Katz on a modified network, where self loops are added, and where non-reciprocal edges are augmented with negative weights. We also give a multilayer interpretation, where negatively weighted walks between layers compensate for backtracking walks on the only non-empty layer. Studying the limit as the attenuation parameter approaches its upper bound also allows us to propose a generalization of eigenvector-based nonbacktracking centrality measure to this directed network setting. In this context, we find that the two-by-two block matrix arising in previous studies focused on undirected networks must be extended to a new three-by-three block structure to allow for directed edges. We illustrate the centrality measure on a synthetic network, where it is shown to eliminate a localization effect present in standard Katz centrality. Finally, we give results for real networks.

*Keywords*: generating function, inverse participation ratio, Katz, localization, loops, multilayer, graph spectrum, backtracking walks.

2000 Math Subject Classification: 34K30, 35K57, 35Q80, 92D25

## 1. Motivation

The tasks of community detection and node centrality measurement can both be addressed by quantifying traversals around a network; either random [21, 28] or deterministic [6, 11]. In particular, the

© The author 2017. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

concept of a walk, which allows nodes and edges to be revisited, forms the basis of the Katz centrality measure [16, 34], whose limiting value corresponds to eigenvector centrality [4, 5]. However, all walks are not created equal. Recently, several authors have argued that certain, backtracking, walks, should be ignored. These walks include at least one back-and-forth flip between a pair of nodes. One concrete justification for the use of nonbacktracking walks is that localization effects can sometimes be avoided [17, 24, 31]. In the context of community detection, a nonbacktracking version of spectral clustering was proposed in [20]. Moreover, quantities defined using the concept of nonbacktracking walks have proved useful in the framework of undirected networks when tackling decycling problems, as well as dismantling problems and the problem of optimal percolation (see, e.g., [25, 26] and references therein).

A nonbacktracking analogue of eigenvector centrality was developed in [24] for undirected networks, and a Katz version was proposed in [14] and studied from a matrix polynomial perspective. However, none of those references handle directed edges. In this work, we therefore address the case of nonbacktracking walk centrality for directed networks.

The paper is organized as follows. In Section 2 we introduce some background concepts and notation and, in particular, set up the task of computing nonbacktracking walk centrality on a directed network. In Section 3 we review a recurrence that counts nonbacktracking walks and leads to a closed form generating function. This allows us to conclude that nonbacktracking walk centrality on a directed network may be computed at the same cost as Katz centrality. Section 4 shows that the new centrality measure may also be interpreted as standard Katz on a modified version of the network, where self loops and negative versions of non-reciprocal edges have been added. We also show that the measure may be viewed from a multilayer perspective. The nonbacktracking walk centrality measure involves a downweighting parameter whose range of values is determined by the radius of convergence of the generating function. We show in Section 5 how this upper limit is related to the spectral radius of a three-by-three block matrix built from the underlying adjacency matrix. We also show that the spectral radius is unchanged when nodes of certain types are removed, and we argue that this property can have computational benefits. Section 6 considers the limiting values of the downweighting parameter. In particular, by allowing the downweighting parameter to approach its upper bound, we arrive at a generalization to directed networks of the nonbacktracking eigenvector centrality measure proposed in [24]. Section 7 gives a rigorous analysis of the new measure on a synthetic network. On this example, in the limit of large network size the nonbacktracking measure can avoid a localization issue present in the Katz version. In Section 8 we give experimental results on some real networks, and we finish in Section 9 with a brief discussion.

## 2. Background and Problem Setting

We begin this section with some definitions and notation. Let $G = (V, E)$ be a directed graph (digraph) with $n$ nodes. Suppose that the graph does not contain self loops (edges from a node to itself) or multiple edges, and assume moreover that it is unweighted, i.e., all its edges have unit weight. The graph can thus be represented by means of an $n \times n$ binary adjacency matrix $A = (a_{ij})$. Here $a_{ij} = 1$ if there is a directed edge from node $i$ to node $j$, and $a_{ij} = 0$ otherwise.

The identity matrix, whose order will be clear from the context, will be denoted by $I$. The $i$th column of the identity matrix of order $n$ will be denoted by $\mathbf{e}_i \in \mathbb{R}^n$. The vector $\mathbf{0} \in \mathbb{R}^n$ will be the vector of all zeros and $\mathbf{1} \in \mathbb{R}^n$ the vector of all ones. We use $\widehat{\mathbf{1}} \in \mathbb{R}^{n-1}$ and $\widehat{\mathbf{e}}_i \in \mathbb{R}^{n-1}$ (for $i = 1, \ldots, n-1$) to denote the vectors containing the first $(n-1)$ components of vectors $\mathbf{1}$ and $\mathbf{e}_i$, respectively. We use $\|\mathbf{v}\|_2$ to denote the Euclidean norm of a vector $\mathbf{v}$, and $\rho(M)$ to denote the spectral radius of a square matrix $M$.

We denote a directed edge from $i$ to $j$ by $i \to j$. We call an edge $i \to j$ *reciprocal*, and denote this by

$i \leftrightarrow j$, if $i \to j \in E$ and $j \to i \in E$. We denote by $S = (s_{ij}) \in \mathbb{R}^{n\times n}$ the adjacency matrix of the subgraph of $G$ obtained by removing all edges that are not reciprocal. The entries of this matrix may thus be defined as $s_{ij} = a_{ij}a_{ji}$.

A *walk of length r* from node $i_1$ to node $i_{r+1}$ is a sequence of $r+1$ nodes $i_1, i_2, \ldots, i_{r+1}$ such that $i_\ell \to i_{\ell+1} \in E$ for all $\ell = 1, 2, \ldots, r$. A standard result shows that $(A^r)_{ij}$ counts the number of walks of length $r$ from node $i$ to node $j$ [10]. We will use the notation

$$a\,b\,c\cdots h$$

to denote a walk that uses nodes $a\,b\,c\cdots h$ in that order. Where necessary, we use $\star$ as a placeholder for a general node and indicate the walk length. So, for example,

$$i \star \ell \star \cdots \star j \qquad \text{of length } r$$

denotes a walk from $i$ to $j$ of length $r$ whose third node must be $\ell$ .

We denote by $D$ the diagonal matrix whose diagonal entries are $D_{ii} = (A^2)_{ii} = (S\mathbf{1})_i$. Hence $D_{ii}$ counts the number of neighbours of $i$ that are connected to $i$ by reciprocal edges. We use $d_i^{\text{out}}$ to denote the *out-degree* of node $i$, that is, the number of nodes that $i$ can reach in one step. So $\mathbf{d}^{\text{out}} := A\mathbf{1}$ is the vector of out-degrees. Similarly, $d_i^{\text{in}}$ denotes the *in-degree* of node $i$, that is the number of nodes that can reach $i$ in one step. In vector form, $\mathbf{d}^{\text{in}} := A^T\mathbf{1}$, where $A^T$ denotes the transpose of $A$.

A walk is said to be *backtracking* if it contains at least one sequence of nodes of the form $i\,\ell\,i$, and *nonbacktracking* otherwise. We will use BTW and NBTW as shorthand for backtracking walk and nonbacktracking walk, respectively. We note that NBTWs have typically been studied on undirected networks [2, 17, 20, 24–26, 32], but the definition continues to make sense in the directed case. We will denote by $p_r(A)$ the matrix whose $(i,j)$th entry counts the number of NBTWs of length $r$ from node $i$ to node $j$. Note that this is the nonbacktracking analogue of the matrix power $A^r$. To illustrate the idea, in Figure 1 we show an example of a windmill network—this class of networks will be studied more comprehensively in Section 7. It is easily verified that

$$p_1(A) = A = \begin{bmatrix} 0&1&0&0&0&0&1\\ 0&0&0&0&0&0&1\\ 0&0&0&1&0&0&1\\ 0&0&0&0&0&0&1\\ 0&0&0&0&0&1&1\\ 0&0&0&0&0&0&1\\ 1&1&1&1&1&1&0 \end{bmatrix}, \quad p_2(A) = A^2 - D = \begin{bmatrix} 0&1&1&1&1&1&1\\ 1&0&1&1&1&1&0\\ 1&1&0&1&1&1&1\\ 1&1&1&0&1&1&0\\ 1&1&1&1&0&1&1\\ 1&1&1&1&1&0&0\\ 0&1&0&1&0&1&0 \end{bmatrix}$$

and

$$p_3(A) = \begin{bmatrix} 1&0&1&2&1&2&0\\ 0&1&0&1&0&1&0\\ 1&2&1&0&1&2&0\\ 0&1&0&1&0&1&0\\ 1&2&1&2&1&0&0\\ 0&1&0&1&0&1&0\\ 0&0&0&0&0&0&3 \end{bmatrix}.$$

For example, there are two walks of length three from node 1 to node 2, given by 1272 and 1712. Both of these are backtracking, and hence $(p_3(A))_{12} = 0$.

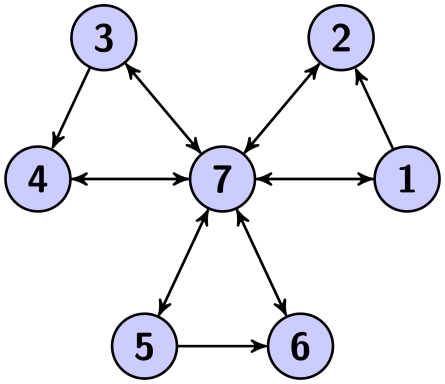


FIG. 1. Windmill network with $m = 3$ triangles.

The concept of walk counting has proved useful in the development of network centrality measures. Katz centrality [16] assigns a value $k_i = \mathbf{e}_i^T \mathbf{k}$ to node $i$ according to

$$(I - \alpha A)\mathbf{k} = \mathbf{1}, \tag{2.1a}$$

where $\alpha > 0$ is a parameter. To ensure that $k_i > 0$, this parameter must be restricted to the range $0 < \alpha < 1/\rho(A)$, in which case we may expand the resolvent $(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots$, so that $k_i$ may be written as

$$k_i = 1 + \sum_{r=1}^{\infty} \sum_{j=1}^{n} \alpha^r (A^r)_{ij}. \tag{2.1b}$$

In this way, we see that $k_i$ from (2.1b) may be viewed as a weighted sum of all walks[1] emanating from node $i$, with the count for walks of length $r$ scaled by $\alpha^r$. In the limit as $\alpha \to 1/\rho(A)$, the Katz centrality vector $\mathbf{k}$ approaches the Perron-Frobenius vector of $A$, which corresponds to eigenvector centrality [4, 5], when the graph under study is strongly connected.

Katz centrality gives the same weight to all walks of the same length. For example, in Figure 1, 717176 has the same influence on the centrality measure as 712756, even though the latter, nonbacktracking, walk, traverses the network more widely and is likely to be more relevant in terms of message passing or disease spreading. This type of argument motivated the authors in [14] to propose and study a nonbacktracking analogue of Katz centrality, restricted to the case of undirected networks, where $A = A^T$. Here, the node centrality vector is defined as

$$\mathbf{b} = \mathbf{1} + \sum_{r=1}^{\infty} t^r p_r(A)\mathbf{1}, \tag{2.2a}$$

where $0 < t < 1$ is a parameter analogous to $\alpha$. Hence, node $i$ is assigned the centrality

$$b_i = 1 + \sum_{r=1}^{\infty} \sum_{j=1}^{n} t^r (p_r(A))_{ij}. \tag{2.2b}$$

[1]The unit shift on the right hand side of (2.1b), which does not affect the relative ordering of the node centralities, may be regarded as arising from a single walk of length zero.

It was shown in [14] that, for undirected networks, the NBTW centrality vector (2.2a) can be computed as cheaply as the Katz version in (2.1a). Further, in an appropriate sense, the limiting value of $\mathbf{b}$ as a function of $t$ in (2.2a) agrees with the nonbacktracking eigenvector centrality measure proposed in [24] for undirected networks.

Our aim here is to consider the NBTW centrality measure (2.2b) in the case where the network has directed edges.

## 3. Recurrence and Generating Function

We now quote a three-term recurrence that relates $p_r(A)$ to $p_{r-1}(A)$, $p_{r-2}(A)$ and $p_{r-3}(A)$, and leads to an expression for the generating function $\sum_{r=0}^{\infty} t^r p_r(A)$. The result may be traced back to [7]. A more accessible treatment from the perspective of graph theory and linear algebra may be found in [33]. We also note that in the case of an undirected network, the recurrence collapses to a two-term expression that was discovered independently in the theory of zeta functions of graphs [32] and exploited in [14] from a network science perspective.

THEOREM 3.1 In the notation above, $p_0(A) = I$, $p_1(A) = A$, $p_2(A) = A^2 - D$ and

$$p_r(A) = Ap_{r-1}(A) + (I-D)\,p_{r-2}(A) - (A-S)\,p_{r-3}(A), \quad \forall r \geqslant 3. \tag{3.1}$$

Further, let $M(t) = I - At + (D-I)t^2 + (A-S)t^3$ and suppose $t$ is such that the power series $\phi(A,t) := \sum_{r=0}^{\infty} t^r p_r(A)$ converges. Then

$$M(t)\phi(A,t) = (1-t^2)I. \tag{3.2}$$

*Proof.* See [7, 33]. □

Recall from (2.2a) that we may compute the NBTW centrality vector as $\mathbf{b} = \phi(A,t)\mathbf{1}$. Theorem 3.1 shows that this may be written as the linear system

$$\left(I - At + (D-I)t^2 + (A-S)t^3\right)\mathbf{b} = (1-t^2)\mathbf{1}. \tag{3.3}$$

We see that the coefficient matrix has the same sparsity as $I - \alpha A$ in (2.1a), which shows that NBTW centrality for a directed network may be computed at least as cheaply as Katz centrality. Further comments on computational complexity will be given in Section 5.

REMARK 3.1 The results in Theorem 3.1 correspond naturally to a recursive proof on walk length where new edges are added at the start of a walk. An alternative recurrence for $p_r(A)$ could be established via a proof that adds edges at the *end* of the walk. This leads to the relationships

$$p_r(A) = p_{r-1}(A)A + p_{r-2}(A)(I-D) - p_{r-3}(A)(A-S)$$

and

$$\phi(A,t)M(t) = (1-t^2)I,$$

which are equivalent to (3.1) and (3.2), respectively.

REMARK 3.2 It is easily checked that equation (3.3) reduces to $(I - At)\mathbf{b} = \mathbf{1}$, i.e., to the linear system (2.1a), when the network does not contain reciprocal edges, and thus $p_r(A) = A^r$.

Let us now define, for $r \geqslant 2$, the $3n \times n$ matrix

$$q_r(A) := \begin{bmatrix} p_r(A) \\ p_{r-1}(A) \\ p_{r-2}(A) \end{bmatrix}. \tag{3.4}$$

Then, if we let

$$C := \begin{bmatrix} A & (I-D) & (S-A) \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \in \mathbb{R}^{3n\times 3n} \tag{3.5}$$

it holds from Theorem 3.1 that $q_r(A) = Cq_{r-1}(A)$ for $r \geqslant 3$, and thus

$$q_r(A) = C^{r-2}q_2(A), \quad \text{for } r \geqslant 2. \tag{3.6}$$

This will prove useful in section 5, where we study the radius of convergence of the power series.

## 4. Other Interpretations

In the previous section we characterized NBTW centrality in terms of a linear system. To add further insight, we now interpret this system heuristically from two alternative viewpoints.

### 4.1 *Katz on a Modified Graph*

To give another interpretation of the NBTW centrality measure in (3.3), we first note that the walk-counting motivation for Katz centrality continues to make sense when the edge weights are nonnegative integers. In this case we interpret an integer weight $s$ as representing $s$ distinct edges between two nodes. For example, in a transport setting, if there are two roads between Pisa and Florence and three roads between Florence and Bologna, then there are $2 \times 3 = 6$ distinct ways to get from Pisa to Bologna with these roads.

Now, the centrality measure $\mathbf{b}$ in (3.3) arises when standard Katz, with $\alpha = t$, is applied to a modified network with adjacency matrix

$$\mathscr{A}(t) := A - (D-I)t - (A-S)t^2,$$

and with the right-hand side rescaled by a factor $1-t^2$. In this modification to the network, each node $i$ is given a self loop with a integer weight, $1-D_{ii}$, scaled by $t$. Also, the unit weight on each nonreciprocal edge $i \to j$ is decreased by $t^2$. It must therefore be the case that adding these self loops and reweighting nonreciprocal edges and then applying the standard Katz procedure, with a rescaled right-hand side, is entirely equivalent to restricting to NBTW counts on the original graph. To understand this equivalence, we may consider what happens when we count "standard walks" by powering up $\mathscr{A}(t)$.

We note that on one hand, this alternative interpretation is conceptually easier, in the sense that we need only consider standard walks. On the other hand, the interpretation is conceptually harder, in the sense that we must relate walks on a modified, weighted graph (with scaled integer weights, that for the loops may even be negative) to NBTWs on the original graph.

When considering traversals around the modified graph, we should regard

- each loop as representing a walk of length two, since the extra factor of $t$ in its weight indicates that it is compensating for traversals that use one extra edge, and

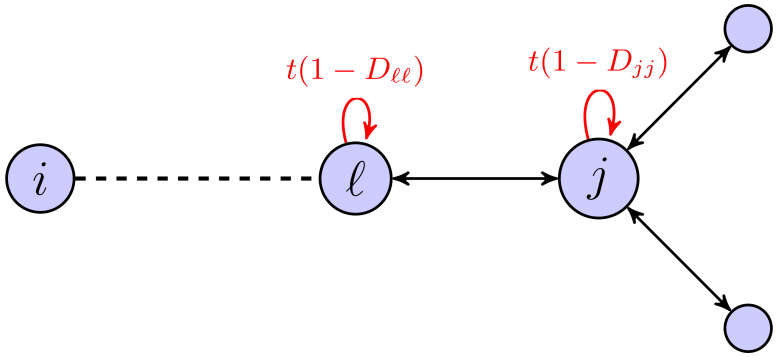


FIG. 2. Illustration of walks on the modified graph, for $\ell \leftrightarrow j$. In reaching node $j$ via its neighbour $\ell$, an immediate backtrack to node $\ell$ is eliminated by the loop at node $\ell$. Hence the continuation that finishes $\ell j \ell j$ must be treated as a special case. Walks finishing $\ell jhj$ for $h \neq \ell$ are eliminated by the loop at node $j$.

- each non-reciprocal edge as representing both a walk of length one and a walk of length three. In the latter case, the weight $t^2$ indicates that the edge is counteracting the presence of traversals that use two extra edges.

Focusing on distinct nodes $i \neq j$ and walks of length $r$, we will give an explanation that relies on induction. So, we assume that the weighted count for NBTWs of any $s < r$ in the original graph is correctly matched by the $t^s$ coefficient in $(1-t^2)\,(I - t\mathscr{A}(t))^{-1}\,\mathbf{1}$. Now, since $\mathscr{A}(t)$ is formed from $A$ by adding extra edges and weights, it is clear that any walk of length $r$ in the original graph is also present in the modified graph, and in both cases this walk contributes to the $t^r$ term in the respective Katz resolvent. Our task is now to show that the modifications in $\mathscr{A}(t)$ reduce the count in such a way that what remains as the $t^r$ coefficient corresponds precisely to the count for NBTWs of length $r$ in $A$.

For convenience, we will form longer walks by adding edges at the *end* of shorter walks. With this viewpoint, the loops in $\mathscr{A}(t)$ are used to balance the presence of BTWs of the form

$$i \star \cdots \star \ell\; j \star j \qquad \text{of length } r.$$

To get an overview of the next step in the argument, we turn to Figures 2 and 3. Here, a walk on the modified graph reaches node $j$ after $r-2$ steps, and it does so via some neighbour $\ell$ of $j$.

**Case 1:** We focus first on the case where the edge $\ell \to j$ is reciprocated, as in Figure 2. Here $s_{\ell j} = s_{j\ell} = 1$ and $(S-A)_{\ell j} = (S-A)_{j\ell} = 0$. The walk may perform a final backtracking double-step either by visiting a distinct neighbour $h \neq \ell$ and returning to $j$, or by revisiting $\ell$ and returning to $j$. We must distinguish between these two types because our inductive argument assumes that we have already removed backtracking walks of length less than $r$—so a walk that revisits node $\ell$ corresponds to a length $r-1$ backtracking walk ending at $\ell$. Such a walk has already been removed from the count, so will not be further propagated when we increase length by one (i.e., multiply by $t\mathscr{A}(t)$). In essence, *the negative loop attached at node $\ell$ has already taken care of this case*. In more detail, ignoring node $\ell$, there are $D_{jj}-1$ distinct neighbours of $j$ such that $j \leftrightarrow \star$, each offering a final backtracking double-step. Their presence is compensated for by the loop at node $j$. This is consistent with $(I - t\mathscr{A}(t))^{-1}\,\mathbf{1}$, giving part

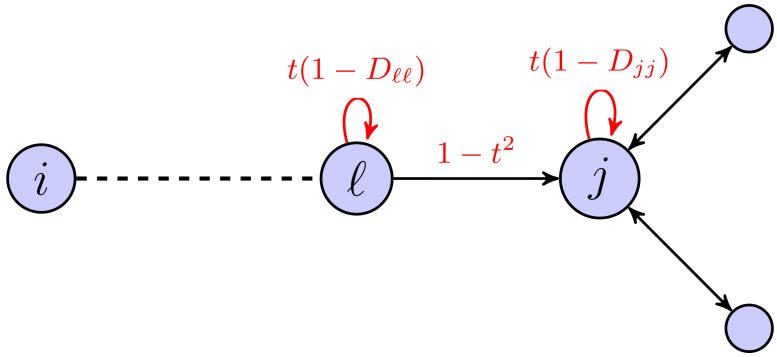


FIG. 3. Illustration of walks on the modified graph, variation of Figure 2 where $\ell \to j$. In this case, a continuation that finishes $\ell j \ell j$ is not possible.

of the required expression. The case

$$i \star \cdots \star \ell\ j\ \ell\ j \qquad \text{of length } r \tag{4.1}$$

must be treated differently because, by induction, walks on the modified graph of the form

$$i \star \cdots \star \ell\ j\ \ell \qquad \text{of length } r-1$$

*have already been removed from the coefficient of* $t^{r-1}$. Hence, we cannot simply propagate the count by adding the final edge $\ell \to j$. Instead, we must propagate the count from length $r-2$ to $r$. The number of walks of the form (4.1) is precisely the number of walks of the form

$$i \star \cdots \star \ell\ j \qquad \text{of length } r-2.$$

Hence, to deal with walks of the form (4.1) we simply subtract the $t^{r-2}$ coefficient. This is consistent with $(I - t\mathscr{A}(t))^{-1}(-t^2)\mathbf{1}$, so, overall, a factor $1-t^2$ arises in the right-hand side of (3.3).

**Case 2:** Now consider the alternative case where the edge $\ell \to j$ is not reciprocated, as in Figure 3. Here $(S-A)_{\ell j} = -1$ and $(S-A)_{j\ell} = 0$. The arguments in Case 1 must be adapted slightly because there is no longer a backtracking walk of length $r-1$ ending at $\ell$. Since this nonexistent walk will not have been removed at the previous level, we must not add it back in, which explains the extra $-(A-S)t^2$ term in $\mathscr{A}(t)$.

### 4.2 *Relationship to Multilayer Networks*

From Theorem 3.1 and (3.6), since $p_0(A) = I$ we see that the NBTW counting matrix $p_r(A)$ may be extracted from the upper left $n \times n$ block of the $r$th power of $C$. The matrix $C$ can be viewed as the supra-adjacency matrix associated with a multilayer network. Here, as shown in Figure 4, we have three layers, each of which contains the same set of nodes. This type of multilayering structure is usually referred to as *node aligned* [18]. In this setting, the diagonal blocks describe interactions within each layer, while the off-diagonal blocks describe interactions between layers. Forming powers of the matrix $C$ counts walks around the multilayer structure where edges within and between layers are treated equally. From (3.5) we see that interactions occurring within layer 1 are described by $A$, while those occurring within layers 2 and 3 are both described by a zero matrix. In particular, having entered layer

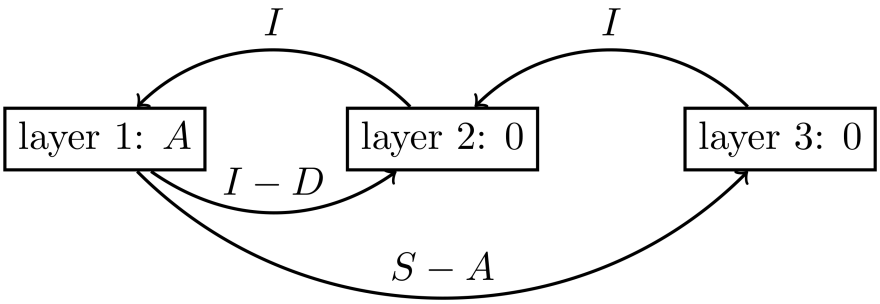


FIG. 4. Representation of the multilayer network described by the matrix $C$ in (3.5).

2 or 3, the walk may be continued only by moving back up to the previous layer. Intuitively, *the effect of moving down the layers is to introduce negatively weighted walks that cancel the contribution of the BTWs within layer 1*. Moving back up the layers, to return to layer 1, causes extra edges to be traversed; so we take one extra step on moving from layer 2 back to layer 1 and two extra steps on moving from layer 3 to layer 2 and then to layer 1. These extra steps are needed to match the number of edges used by the equivalent BTWs taking place within layer 1.

## 5. The Radius of Convergence

In this section we study the radius of convergence of the power series $\phi(A,t)$ in Theorem 3.1. The parameter $t > 0$ represents a weight for walks taking place in the network under study; therefore, it is natural to require $t \in (0,1)$, so that shorter walks are given more importance than longer ones.[2] From a practical perspective, it is necessary to know what further restriction, if any, is required. The following theorem shows that $1/\rho(C)$ is a lower bound on the radius of convergence which is sharp when there is at least one directed cycle, where $C$ is defined in in (3.5). An example, where there are no directed cycles is given by

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Here the radius of convergence for $\phi(A,t)$ is infinite; whereas $1/\rho(C) = 1$.

THEOREM 5.1 The radius of convergence $R_\phi$ of the power series $\phi(A,t) = \sum_{r=0}^{\infty} t^r p_r(A)$ is

$$R_\phi = \begin{cases} 1/\rho(C) & \text{if there is at least one directed cycle;} \\ \infty & \text{if there are no directed cycles, in which case } \rho(C) = 1. \end{cases}$$

*Proof.* Recall (3.4) and (3.6), and, for the sake of convenience, set $p_r(A) = 0$ for all $r < 0$. We then

[2]It is worth mentioning that the series defining $\phi(A,t)$ makes sense for $t \in \mathbb{C}$. We restrict ourselves to the case of $t \in (0,1)$ since we are interested in the application to network science.

have

$$q_1(A) := \begin{bmatrix} A \\ I \\ 0 \end{bmatrix} \quad \text{and} \quad q_0(A) := \begin{bmatrix} I \\ 0 \\ 0 \end{bmatrix},$$

and thus $q_1(A) = Cq_0(A)$ and $q_2(A) = Cq_1(A) - q_0(A) = (C^2 - I)q_0(A)$. Then a straightforward computation shows

$$\begin{aligned}
\phi(A,t) = \sum_{r=0}^{\infty} t^r p_r(A) &= [I,0,0]^T \sum_{r=2}^{\infty} t^r q_r(A) + I + tA \\
&= q_0(A)^T \left( \sum_{r=2}^{\infty} t^r C^{r-2} q_2(A) \right) + I + tA \\
&= q_0(A)^T \left( \sum_{r=2}^{\infty} t^r C^{r-2} (C^2 - I) q_0(A) + q_0(A) + t q_1(A) \right) \\
&= q_0(A)^T \left( \sum_{r=2}^{\infty} t^r C^r q_0(A) + q_0(A) + tCq_0(A) - t^2 \sum_{r=0}^{\infty} t^r C^r q_0(A) \right) \\
&= q_0(A)^T \left( (1-t^2) \sum_{r=0}^{\infty} t^r C^r \right) q_0(A)
\end{aligned}$$

and thus the series $\phi(A,t)$ converges if $\sum_{r=0}^{\infty} t^r C^r$ does, i.e., if $t\rho(C) < 1$. Moreover, the only possibility for $R_\phi \neq 1/\rho(C)$ is when $\rho(C) = 1$.

Suppose that there are no directed cycles. Then, there are no NBTWs of arbitrary length, and hence $R_\phi = \infty$.

Conversely, suppose that there is at least one directed cycle. Note first that 1 is an eigenvalue of $C$, since $C\mathbf{1} = \mathbf{1}$, and hence, $\rho(C) \geqslant 1$. There are two cases. If $\rho(C) > 1$, then the first part of the proof shows that $R_\phi = 1/\rho(C)$. If $\rho(C) = 1$, observe if $i$ is a node in a directed cycle of length $k$ then $(p_{nk}(A))_{ii} \geqslant 1$ for all $n = 1,2,3,\ldots$, implying $R_\phi \leqslant 1$. Hence, $1 = 1/\rho(C) \leqslant R_\phi \leqslant 1$, and thus $R_\phi = 1 = 1/\rho(C)$. □

THEOREM 5.2 Let $R_\psi$ be the radius of convergence of $\psi(A,t) = \sum_{r=0}^{\infty} t^r A^r$ and let $R_\phi$ be the radius of convergence of $\phi(A,t) = \sum_{r=0}^{\infty} t^r p_r(A)$. Then, $R_\psi \leqslant R_\phi$.

*Proof.* Since every nonbacktracking walk is a walk, elementwise it holds that $p_r(A) \leqslant A^r$. This implies that elementwise

$$|\phi(A,t)| \leqslant |\psi(A,t)|,$$

and hence the radius of convergence of the right-hand side is less than or equal to the radius of convergence of the left-hand side. □

Theorem 5.2 shows that the interval in which the downweighting parameter of Katz centrality is allowed to vary is not reduced when $f(z) = (1-tz)^{-1}$ is applied to $\mathscr{A}(t)$ instead of $A$. Indeed, generically, in the former case $t \in (0, 1/\rho(C))$, while in the latter $t \in (0, 1/\rho(A))$.

In the next subsection, we show that the spectral radius of $C$ does not change when certain types of nodes are removed from the graph. This property is not shared by the matrix $A$, whose spectrum changes, in general, under this type of node (and edge) removal.

### 5.1 *Pruning and the Spectral Radius*

In the remainder of this section we assume without loss of generality that the graph under study is weakly connected, i.e., that each node in the network is reachable from any other node when all the edges are regarded as having no orientation. (Otherwise, the problems that we study below break down into essentially independent subtasks. A node $i$ will be called a *source node* if it only has outgoing links and thus $d_i^{\text{out}} > 0$ and $d_i^{\text{in}} = 0$; similarly, $i$ will be referred to as a *dangling node* if $d_i^{\text{in}} > 0$ and $d_i^{\text{out}} = 0$, i.e., if it only has incoming links. For the sake of clarity in the exposition, we will assume that the graph represented by $A$ has only one source (resp., dangling) node, namely node $n$.[3] If the graph has more than one, the following reasoning can be applied iteratively. In the remainder of this section we adopt the following notation. We denote by $\widehat{A} = (\widehat{a}_{ij})$ the $(n-1)\times(n-1)$ adjacency matrix of the graph obtained after node $n$ is removed. The matrix $\widehat{S} = (\widehat{s}_{ij})$ will have entries $\widehat{s}_{ij} = \widehat{a}_{ij}\widehat{a}_{ji}$ and the diagonal matrix $\widehat{D}$ will be such that $\widehat{D}_{ii} = (\widehat{A}^2)_{ii}$. If $n$ is a source, then $A\mathbf{e}_n = \mathbf{0}$ and $\mathbf{e}_n^T A = [\mathbf{a}_R^T, 0] = [a_{n1}, a_{n2}, \ldots, a_{n,n-1}, 0]$, where $A = (a_{ij})$ are the entries of the adjacency matrix; similarly, if $n$ is a dangling node, then $\mathbf{e}_n^T A = \mathbf{0}^T$ and $A\mathbf{e}_n = [\mathbf{a}_C^T, 0]^T = [a_{1n}, a_{2n}, \ldots, a_{n-1,n}, 0]^T$. Note that, since the graph is assumed to be weakly connected and since $n$ is either a source or a dangling node, $\mathbf{a}_C \neq \widehat{\mathbf{0}}$ or $\mathbf{a}_R \neq \widehat{\mathbf{0}}$. Once node $n$ is removed from the graph, the three-by-three block matrix analogue of (3.5) used to determine the number of NBTWs in the new graph has the form

$$\widehat{C} := \begin{bmatrix} \widehat{A} & (I-\widehat{D}) & (\widehat{S}-\widehat{A}) \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \in \mathbb{R}^{(3n-3)\times(3n-3)}.$$

The following theorem tells us that removing source and dangling nodes from a network does not alter the spectral radius of the three-by-three block matrix used to determine the number of NBTWs.

THEOREM 5.3 In the above notation,

$$\det(\lambda I - C) = (\lambda^3 - \lambda)\det(\lambda I - \widehat{C}).$$

*Proof.* Let $F(\lambda) = I\lambda^3 - A\lambda^2 + (D-I)\lambda + (A-S)$. The matrix equation

$$\begin{bmatrix} \lambda I - A & D-I & A-S \\ -I & \lambda I & 0 \\ 0 & -I & \lambda I \end{bmatrix} \begin{bmatrix} I & \lambda I & \lambda^2 I \\ 0 & I & \lambda I \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} \star & \star & F(\lambda) \\ -I & 0 & 0 \\ 0 & -I & 0 \end{bmatrix},$$

where $\star$ denotes a block whose exact form is not relevant for the argument, shows that $\det(\lambda I - C) = \det F(\lambda)$. Similarly, one can prove that $\det(\lambda I - \widehat{C}) = \det \widehat{F}(\lambda) =: \det(I\lambda^3 - \widehat{A}\lambda^2 + (\widehat{D}-I)\lambda + (\widehat{A}-\widehat{S}))$.

It is easily seen that

$$F(\lambda) = \begin{bmatrix} \widehat{F}(\lambda) & g(\lambda)\mathbf{a}_C \\ h(\lambda)\mathbf{a}_R^T & \lambda^3 - \lambda \end{bmatrix}, \tag{5.1}$$

where the functions $g(\lambda)$ and $h(\lambda)$ are defined as:

(1) $g(\lambda) = 0$, $h(\lambda) = 1 - \lambda^2$ if $d_n^{\text{out}} = 0$;

[3] This can be obtained after relabelling the nodes and, consequently, after a symmetric permutation applied to the matrix $A$.

(2) $h(\lambda) = 0$, $g(\lambda) = 1 - \lambda^2$ if $d_n^{\text{in}} = 0$.

Hence, $F(\lambda)$ is block triangular and the statement immediately follows. □

The class of nodes that can be removed from the network without affecting the radius of convergence of $\phi(A,t) = \sum_r p_r(A)t^r$ is broader than the one we just described. Let us call node $i$ a *reciprocal leaf* connected to node $j$ if $A\mathbf{e}_i = A^T\mathbf{e}_i = \mathbf{e}_j$; here, the node has a single neighbour, connected by a reciprocal edge. Suppose now with no loss of generality that node $n$ is a reciprocal leaf and is connected to node $n-1$. The following result shows that we can also safely remove reciprocal leaves from the network.

THEOREM 5.4 In the notation above,

$$\det(\lambda I - C) = \lambda^3 \det(\lambda I - \widehat{C}).$$

*Proof.* A proof may be given along the same lines of that of Theorem 5.3 by using the fact that now

$$F(\lambda) = \begin{bmatrix} \widehat{F}(\lambda) + \lambda \widehat{\mathbf{e}}_{n-1}\widehat{\mathbf{e}}_{n-1}^T & -\lambda^2 \widehat{\mathbf{e}}_{n-1} \\ -\lambda^2 \widehat{\mathbf{e}}_{n-1}^T & \lambda^3 \end{bmatrix}$$

and

$$\begin{aligned} \det F(\lambda) &= \lambda^3 \det(\widehat{F}(\lambda) + \lambda \widehat{\mathbf{e}}_{n-1}\widehat{\mathbf{e}}_{n-1}^T) + \lambda^2 \det\left( \begin{bmatrix} (\widehat{F}(\lambda))_{1:n-2,1:n-1} \\ -\lambda^2 \widehat{\mathbf{e}}_{n-1}^T \end{bmatrix} \right) \\ &= \lambda^3 \det(\widehat{F}(\lambda) + \lambda \widehat{\mathbf{e}}_{n-1}\widehat{\mathbf{e}}_{n-1}^T) - \lambda^4 \det(\widehat{F}(\lambda))_{1:n-2,1:n-2} \\ &= \lambda^3 \det \widehat{F}(\lambda) + \lambda^4 \det(\widehat{F}(\lambda))_{1:n-2,1:n-2} - \lambda^4 \det(\widehat{F}(\lambda))_{1:n-2,1:n-2} \\ &= \lambda^3 \det \widehat{F}(\lambda), \end{aligned}$$

where in the second-to-last equality we have expanded the determinant with respect to the last row (or, equivalently, column). □

Theorems 5.3 and 5.4 are computationally relevant because, as established in Theorem 5.1, in order to understand what range of $t$ values is valid, we require $\rho(C)$. The results show that before computing $\rho(C)$ we are allowed to iteratively prune sources, dangling nodes and reciprocal leaves from the network. In the next section, we show for completeness that the linear system defining the NBTW centralities is also amenable to the same type of pruning. We note that similar ideas have been proposed to reduce the cost of the linear system solve asssociated with Katz and PageRank, where the coefficient matrix has a related structure; see, for example, [3, 22].

### 5.2 *Pruning the linear system*

In this section we study how the linear system (3.3) changes when source nodes, dangling nodes and reciprocal leaves are removed from the network. First, let us assume, without loss of generality, that node $n$ is the only source (resp., dangling) node in the network. We describe what happens in the two different cases, depending on the type of node. We adopt the same notation as in subsection 5.1. Furthermore, $\widehat{\mathbf{b}}$ will denote the vector containing the first $n-1$ components of the NBTW centrality vector $\mathbf{b} = (b_i)$.

(1) $d_n^{\text{in}} = 0$. Intuitively, since $n$ is a source node, it sends information to the rest of the graph, and thus its importance, as measured in terms of walk counts, will depend on the importance of the nodes it is

pointing to. On the other hand, the importance of the other nodes will not be affected by the importance of node $n$, since there is no route connecting these nodes to $n$. Algebraically, the matrix $M(t)$ can be written as

$$M(t) = \begin{bmatrix} \widehat{M}(t) & \mathbf{0} \\ (t^3 - t)\mathbf{a}_R^T & 1 - t^2 \end{bmatrix},$$

where here and in the following $\widehat{M}(t) = I - \widehat{A}t + (\widehat{D} - I)t^2 + (\widehat{A} - \widehat{S})t^3$, and thus (3.3) is equivalent to

$$\begin{cases} \widehat{M}(t)\widehat{\mathbf{b}} = (1 - t^2)\widehat{\mathbf{1}} \\ b_n = 1 + t\mathbf{a}_R^T\widehat{\mathbf{b}}. \end{cases}$$

(2) $d_n^{\text{out}} = 0$. Here, because $n$ has no access to the rest of the network, we know that its centrality value is 1. Indeed, the linear system (3.3) can be rewritten

$$\begin{cases} \widehat{M}(t)\widehat{\mathbf{b}} = (1 - t^2)[\widehat{\mathbf{1}} + t\mathbf{a}_C] \\ b_n = 1. \end{cases}$$

Note that the nodes pointing to node $n$ and, more generally, the other nodes that can reach it, gain some of their importance from the presence of node $n$. This is why the right-hand side of the system is perturbed by the pruning operation.

Let us now consider the case where $n$ is a reciprocal leaf connected to node $n-1$. This last case results as a combination of the above two, when $\mathbf{a}_C = \mathbf{a}_R = \widehat{\mathbf{e}}_{n-1}$; the importance of node $n$ is non-trivial and that of all the other nodes is computed by modifying the right-hand side of the system to take into account the leaf removal. It can be shown that

$$M(t) = \begin{bmatrix} \widehat{M}(t) + t^2\widehat{\mathbf{e}}_{n-1}\widehat{\mathbf{e}}_{n-1}^T & -t\widehat{\mathbf{e}}_{n-1} \\ -t\widehat{\mathbf{e}}_{n-1}^T & 1 \end{bmatrix}$$

and hence (3.3) can be written as

$$\begin{cases} \widehat{M}(t)\widehat{\mathbf{b}} = (1 - t^2)[\widehat{\mathbf{1}} + t\widehat{\mathbf{e}}_{n-1}] \\ b_n = (1 - t^2) + tb_{n-1}. \end{cases}$$

Overall, we conclude that the NBTW measure may be computed by solving a linear system *whose dimension is that of the network remaining after all source/dangling nodes and reciprocal leaves have been pruned*, and then, if required, recovering the centrality of the pruned nodes via simple scalar recurrences.

## 6. Limiting Behavior

In [14] it was shown in the undirected case that the Katz-style NBTW measure recovers the eigenvector version of [24] as the attenuation parameter approaches its upper limit. Here we show that in the more general case of directed edges, a related eigenvector limit arises. This gives a natural generalization of the work in [24] to the directed case, and we note that the lack of symmetry takes us from a two-by-two block matrix to a new three-by-three extension.

In order to make the dependence on the attenuation parameter explicit, in the remainder of this section we use $\mathbf{b}(t)$ in place of $\mathbf{b}$ to denote the NBTW centrality vector. The theorem also considers the more straightforward $t \to 0^+$ limit.

THEOREM 6.1 Let $\mathbf{b}(t)$ be the NBTW centrality vector, defined as in (3.3), with $t \in (0, \rho(C)^{-1})$. Then, the ranking produced by $\mathbf{b}(t)$ converges to that produced by the vector of outdegrees $\mathbf{d}^{\text{out}} = (d_i^{\text{out}})$ as $t \to 0^+$.

Moreover, suppose that $I - \rho(C)^{-1}C$ has rank $3n-1$. Then, as $t \to (1/\rho(C))^-$ the ranking produced by $\mathbf{b}(t)$ converges to that given by the first $n$ components of the eigenvector of $C$ associated with the eigenvalue $\rho(C)$.

*Proof.* By using (2.2b) and the definition of $\phi(A,t)$ one finds that entrywise

$$b_i(t) = \mathbf{e}_i^T \phi(A,t)\mathbf{1} = 1 + d_i^{\text{out}} t + O(t^2)$$

and the first part of the statement follows.

For the second part, we first note that $\rho(C)$ must be an eigenvalue of $C$.[4] Then note that, since $\mathbf{b}(t)$ solves (3.3), it follows that

$$(I - tC)\begin{bmatrix} \mathbf{b}(t) \\ t\mathbf{b}(t) \\ t^2\mathbf{b}(t) \end{bmatrix} = \begin{bmatrix} (I - At + (D-I)t^2 + (A-S)t^3)\mathbf{b}(t) \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Hence, for $0 < t < \rho(C)^{-1}$, $\mathbf{b}(t)$ solves the NBTW system (3.3) if and only if

$$(I - tC)\begin{bmatrix} \mathbf{b}(t) \\ t\mathbf{b}(t) \\ t^2\mathbf{b}(t) \end{bmatrix} = (1-t^2)\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

So, to study the limiting behaviour, $t \to \rho(C)$, we can focus on

$$(1-t^2)(I - tC)^{-1}\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

At this point observe that $I - tC$ can be seen as a matrix whose entries depend analytically on the real variable $t$. As such, it admits an analytic singular value decomposition (see, e.g., [8, 15, 29]). In more detail, we can write

$$I - tC = U(t)\Sigma(t)V(t)^T = \sum_{i=1}^{3n} \sigma_i(t)\mathbf{u}_i(t)\mathbf{v}_i(t),$$

where $U(t), V(t)$ are orthogonal and analytic for all real $t$, $\Sigma(t)$ is diagonal and analytic for all real $t$, $\mathbf{u}_i(t)$ (resp., $\mathbf{v}_i(t)$) is the $i$th column of $U(t)$ (resp., $V(t)$) and $\sigma_i(t) = (\Sigma(t))_{ii}$. We assume, without loss of generality, that the functions $\sigma_i(t)$ are labelled in such a way that

$$\sigma_1(\rho(C)^{-1}) \geqslant \cdots \geqslant \sigma_{3n-1}(\rho(C)^{-1}) > \sigma_{3n}(\rho(C)^{-1}) = 0.$$

Since $I - tC$ is invertible for all $0 \leqslant t < \rho(C)^{-1}$, we can write

$$(1-t^2)(I - tC)^{-1}\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = (1-t^2)\sum_{i=1}^{3n} \frac{\upsilon_i(t)}{\sigma_i(t)}\mathbf{u}_i(t) = (1-t^2)\frac{\upsilon_{3n}(t)}{\sigma_{3n}(t)}\mathbf{u}_{3n}(t) + O\left(\frac{\sigma_{3n}(t)}{\sigma_{3n-1}(t)}\right),$$

[4] This follows from results in the theory of matrix polynomials, see, e.g., [12, 14, 23, 27, 30] and references therein.

where

$$\upsilon_i(t) := \mathbf{v}_i(t)^T \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Now, define

$$\widetilde{\mathbf{b}}(t) = \frac{\sigma_{3n}(t)}{(1-t)^2 \upsilon_{3n}(t)} \mathbf{b}(t).$$

The rankings given by $\widetilde{\mathbf{b}}(t)$ and $\mathbf{b}(t)$ are equivalent, since they only differ by an overall scalar multiplicative factor.

To conclude the proof, it now suffices to observe that

- $\upsilon_{3n}(t) \neq 0$ for all $t \in [0, \rho(C)^{-1}]$, as can be shown by technical arguments similar to those in [14];
- $\sigma_{3n}(t)/\sigma_{3n-1}(t) \to 0$ for $t \to \rho(C)^{-1}$, since by assumption $I - \rho(C)^{-1}C$ has rank $3n-1$;
- $\mathbf{u}_{3n}(t)$ is the right singular vector corresponding to $\sigma_{3n}(t)$, and therefore converges to the right null space of $I - \rho(C)^{-1}C$ as $t \to \rho(C)^{-1}$, i.e., the eigenvector of $C$ associated with $\rho(C)$.

□

We note that, using the results in [32, 33], it is possible to prove that the assumption that $(I - \rho(C)^{-1}C)$ has rank $n-1$ is automatically satisfied when the underlying graph is strongly connected. Hence, any problem may be broken down into a set of subproblems for which the assumption holds.

## 7. Analysis for Windmill Networks

In this section, we introduce and study windmill networks; the case with $m = 3$ triangles was defined in Section 2 and illustrated in Figure 1. Our aim is to give a rigorous analysis of a large-scale directed network example where Katz and NBTW centrality can be shown to behave differently. In both cases the generating function has a finite radius of convergence. However, the restriction required for Katz centrality produces a localization effect that can be avoided with the NBTW version.

To quantify localization, we use the approach of [13]. Suppose we have a family of unit Euclidean norm vectors $\mathbf{x} \in \mathbb{R}^n$, defined for all large $n$. Then the *inverse participation ratio* is defined to be $\mathscr{S}(\mathbf{x}) := \sum_{i=1}^n x_i^4$. The family of vectors is said to be *localized* if $\mathscr{S} = O(1)$ and *nonlocalized* if $\mathscr{S} = o(1)$, as $n \to \infty$. Intuitively, localization implies that the majority of the mass in the vector is confined to a finite subset of components (in our case a single component).

Figure 1 illustrates a directed windmill network with $m = 3$ triangles. In the general case we have a hub that is involved in $m$ triangles. The network has $n = 2m+1$ nodes. Labelling the hub as node $n$ for convenience, we have the following edges

- undirected edges from node $n$ to all other nodes in the network,
- a directed edge from node $2i-1$ to node $2i$ for all $1 \leqslant i \leqslant m$.

By symmetry, when we study an eigenvector of $A$ or $C$, or a vector of walk-based centrality measures, on this class of network, there will be at most three distinct components; these will correspond to the odd-numbered nodes from 1 to $2m-1$, the even-numbered nodes, and the hub node $n$. Hence, we will use nodes 1, 2 and $n$ as representatives.

Before giving a full asymptotic analysis, we will make some general remarks. For small $\alpha$ or $t$, the Katz and NBTW centralities correspond to out-degree centrality. Since node $n$ has out-degree $2m$ whereas the remaining nodes have out-degrees of either one or two, it follows that, with out-degree centrality, node $n$ is assigned almost all the measure: $d_n^{\text{out}}/\|\mathbf{d}^{\text{out}}\|_2 = 1 + O(m^{-1})$, and $d_i^{\text{out}}/\|\mathbf{d}^{\text{out}}\|_2 = O(m^{-1})$ for all other $i$. Hence, the inverse participation ratio is $\mathscr{S}(\mathbf{d}^{\text{out}}/\|\mathbf{d}^{\text{out}}\|_2) = 1 + O(m^{-1})$ and the measure is localized. We show below that Katz centrality does not allow $\alpha$ to become sufficiently large to overcome this effect, whereas NBTW centrality does. Referring to Figure 1 we see that if we count standard walks starting from node $n$, we may backtrack to node $n$ on every second step, e.g., after using the edge $7 \to 1$, we may return to 7 via $1 \to 7$, whence we have a further $2m$ choices for the next edge. Hence, the overall walk count can *grow by a factor* $2m$ *every two steps*. If we do not allow backtracking, then a full, *one-way*, triangle must be completed before we can return to the hub; so the overall walk count can only *grow by a factor of roughly* $m$ *every three steps*. This relates to the terms $(2m)^{\frac{1}{2}}$ and $m^{\frac{1}{3}}$ that arise in the analysis below.

### 7.1 *Katz Centrality*

To study Katz centrality on this example, we first consider the spectrum of the adjacency matrix $A$. An eigenpair satisfies $A\mathbf{x} = \lambda\mathbf{x}$, which reduces to

$$mx_1 + mx_2 = \lambda x_n, \quad x_2 + x_n = \lambda x_1, \quad x_n = \lambda x_2.$$

Requiring $\mathbf{x} \neq \mathbf{0}$ leads to the condition

$$\lambda^3 - 2m\lambda - m = 0. \tag{7.1}$$

Straightforward analysis shows that, for large $m$, this cubic has three real roots. The largest root in modulus, $\lambda_1$, is positive and bounded below by $\sqrt{2m}$. Moreover, given any $\varepsilon > 0$ there exists $m_0 \in \mathbb{N}$ such that $m > m_0 \Rightarrow \sqrt{2m} < \lambda_1 < \sqrt{2m} + \varepsilon$. Therefore, taking $0 < \alpha < 1/\sqrt{2m}$ is a sufficient condition for the Katz parameter to be within the allowed range, and it is a good approximation of the exact allowed interval for large values of $m$.

Now, the Katz system (2.1a) reduces to

$$k_n - \alpha(mk_1 + mk_2) = 1, \quad k_1 - \alpha(k_2 + k_n) = 1, \quad k_2 - \alpha k_n = 1.$$

This solves to give

$$k_n = \frac{1 + 2m\alpha + m\alpha^2}{1 - 2m\alpha^2 - m\alpha^3}, \quad k_2 = \frac{1 + \alpha}{1 - 2m\alpha^2 - m\alpha^3}, \quad k_1 = \frac{1 + 2\alpha + \alpha^2}{1 - 2m\alpha^2 - m\alpha^3}. \tag{7.2}$$

We note that as $\alpha$ increases away from zero, the denominator in these expressions vanishes precisely at the reciprocal of the largest root of the cubic appearing in (7.1).

To choose a large $\alpha$ that is consistent with the upper bound, we set $\alpha = c/\sqrt{2m}$, for some $0 < c < 1$. We are interested in the behaviour of the inverse participation ratio in the limit $m \to \infty$. Using standard asymptotic expansions, we find that the the normalized Katz vector, $\widetilde{\mathbf{k}} := \mathbf{k}/\|\mathbf{k}\|_2$ satisfies

$$\widetilde{k}_n = \frac{c}{\sqrt{1 + c^2}}\left(1 + O(1/\sqrt{m})\right), \qquad \widetilde{k}_2 = O(1/\sqrt{m}), \qquad \widetilde{k}_1 = O(1/\sqrt{m}).$$

The inverse participation ratio for Katz centrality is then found to be

$$\mathscr{S}(\widetilde{\mathbf{k}}) = \sum_{i=1}^{n} \widetilde{k}_i^4 = \frac{c^4}{(1 + c^2)^2} + O(1/\sqrt{m}).$$

We see that this quantity is $O(1)$, and hence the centrality measure is localized.

### 7.2 *Nonbacktracking Walk Centrality*

To investigate NBTW centrality on a windmill network, we begin by characterizing the spectral radius of the matrix $C$ in (3.5). Write $C\mathbf{v} = \lambda\mathbf{v}$, where $\lambda \neq 0$, and partition $\mathbf{v}$ into three vectors of equal dimension: $\mathbf{v}^T = [\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T]$. Then, after eliminating $\mathbf{y}$ and $\mathbf{z}$, we arrive at

$$\left(\lambda^3 I - \lambda^2 A + \lambda(D-I) + (A-S)\right)\mathbf{x} = \mathbf{0}.$$

The relevant equations for $x_n$, $x_1$ and $x_2$, are then, respectively,

$$\begin{aligned} \lambda^3 x_n - \lambda^2(mx_1 + mx_2) + \lambda(2m-1)x_n &= 0, \\ \lambda^3 x_1 - \lambda^2(x_2 + x_n) + x_2 &= 0, \\ \lambda^3 x_2 - \lambda^2 x_n &= 0. \end{aligned}$$

Eliminating $x_1$ and $x_2$, and using $\lambda \neq 0$ and $x_n \neq 0$, we arrive at

$$\lambda^5 - \lambda^3 - m\lambda^2 + m = 0.$$

The roots are $\pm 1$ and the cube roots of $m$. So $\rho(C) = m^{1/3}$.

The directed NBTW linear system (3.3) solves to give

$$\begin{aligned} b_n &= \frac{1 + 2mt + (m-1)t^2 - 2mt^3 - 2mt^4 + mt^6}{1 - t^2 - mt^3 + mt^5}, \\ b_2 &= \frac{1 + t + (2m-2)t^2 - t^3 - (2m-1)t^4}{1 - t^2 - mt^3 + mt^5}, \\ b_1 &= \frac{1 + 2t + (2m-1)t^2 + (2m-4)t^3 - (2m+1)t^4 - (4m-3)t^5 + t^6 + (2m-1)t^7}{1 - t^2 - mt^3 + mt^5}. \end{aligned}$$

We note that in each case the denominator has roots $\pm 1$ and the cube roots of $1/m$. Hence, as $t$ increases from zero the denominators first vanish at $t = m^{-1/3}$, which is the reciprocal of $\rho(C)$. This is consistent with Theorem 5.1. Setting $t = cm^{-\frac{1}{3}}$, for some $0 < c < 1$, we find that the normalized NBTW centrality vector $\widetilde{\mathbf{b}} := \mathbf{b}/\|\,\mathbf{b}\,\|_2$ has components

$$\widetilde{b}_n = \frac{m^{-\frac{1}{6}}}{\sqrt{2}c}\left(1 + O(m^{-\frac{1}{3}})\right), \quad \widetilde{b}_2 = \frac{m^{-\frac{1}{2}}}{\sqrt{2}}\left(1 + O(m^{-\frac{1}{3}})\right), \quad \widetilde{b}_1 = \frac{m^{-\frac{1}{2}}}{\sqrt{2}}\left(1 + O(m^{-\frac{1}{3}})\right),$$

and the inverse participation ratio has the expansion

$$\mathscr{S}(\widetilde{\mathbf{b}}) = \sum_{i=1}^{n} \widetilde{b}_i^4 = \frac{m^{-\frac{2}{3}}}{4c^4} + O(m^{-1}).$$

We see that with this choice of $t$, the NBTW centrality measure is nonlocalized.

## 8. Experiments

In this section we describe the results of some numerical tests performed on real-world networks. All digraphs used here can be found in the SuiteSparse Matrix Collection, formerly known as the University of Florida Sparse Matrix Collection, [9]. The network PAJEK/GLOSSGT represents connections between words from the graph/digraph glossary. PAJEK/CALIFORNIA was constructed by expanding a 200-page response set to a search engine query 'California', as used for the HITS algorithm [19]. The network GLEICH/WB-CS-STANFORD reflects the Stanford CS web, where a link in the graph represents an actual link in the web. This network has 1299 nonzeros on the main diagonal which we removed before performing our computations. The last three networks all belong to the SNAP group. SNAP/EPINIONS is a who-trusts-whom online social network of a general consumer review site: `Epinions.com`. The network SNAP/WIKI-TALK summarises users and discussions from the inception of Wikipedia until January 2008. Nodes in the network represent Wikipedia users and a directed edge from node $i$ to node $j$ indicates that user $i$ at least once edited a talk page of user $j$. Finally, SNAP/CIT-PATENTS is a citation network among US Patents. The data set includes all the utility patents granted between 1975 and 1999 and all citations made by them.

### 8.1 *NBTW Centrality vs. Katz*

We first show with a small example how NBTW broadcast centrality may differ from Katz centrality when identifying the most important broadcasters in a network. We performed our test on the small network PAJEK/GLOSSGT. This network contains 72 nodes, but we restrict ourselves to its largest weakly connected component, which has 60 nodes. For this digraph it holds that $\rho(A) = \rho(C) = 1$ and thus the parameters used in the computation of the two centrality indices can be selected to be the same. In our test $\alpha = t = 0.9$. Figure 5 scatter plots the NBTW and Katz centrality vectors. It can be seen that the two centrality measures do not identify the same top broadcasters. Moreover, looking at the correlation coefficients between the rankings, we have that the Pearson's correlation coefficient between the Katz and NBTW centralities is 0.49, so there is not a strong correlation.

In Figures 6 and 7 we display two network visualisations, where node size is proportional to centrality; with Katz in Figure 6 and NBTW in Figure 7. Two nodes share the top Katz centrality. These are connected through a reciprocal edge and are pointed to by the third Katz-ranked node, which is a source node with outdegree equal to one. It can thus be argued that these nodes are identified as the most influential because they point to or are involved in a closed cycle of small length. Indeed, when computing the Katz centrality, short, backtracking walks, are given as much importance as nonbacktracking walks of the same length. In this example, the NBTW measure gives top ranking to a node with outdegree two and, as is also clear from Figure 5, generally provides a more even (less localized) spread of values across the nodes.

### 8.2 *Nodal Pruning*

We saw in subsections 5.1 and 5.2 that both the cost of computing $\rho(C)$ in order to establish a suitable range for the parameter $t$ and the cost of solving the overall linear system (3.3) may be reduced by iteratively removing source/dangling nodes and reciprocal leaves until none remain. The benefits will, of course, depend on the level of dimension reduction achieved. To investigate this issue, in this set of

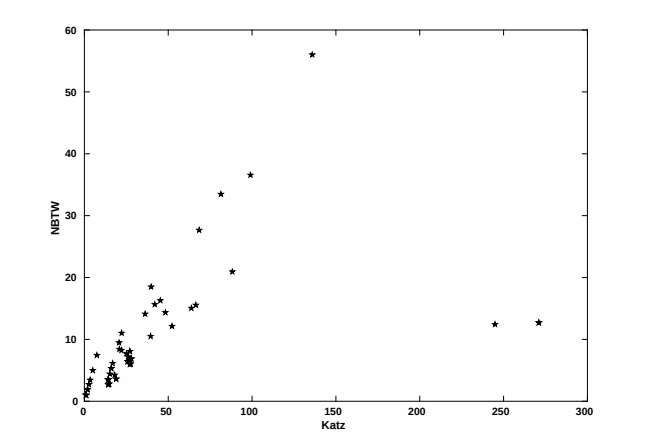


FIG. 5. Scatter plot of the Katz (horizontal axis) and NBTW (vertical axis) centrality vectors. Note that the top Katz centrality score is shared by two nodes, which also have identical NBTW centrality; in this case the two symbols occupy the same place in the scatter plot.

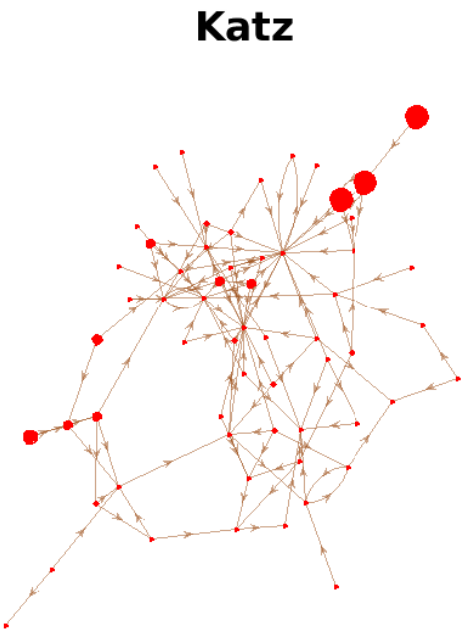


FIG. 6. Representation of the network PAJEK/GLOSSGT where node size is proportional to Katz centrality.

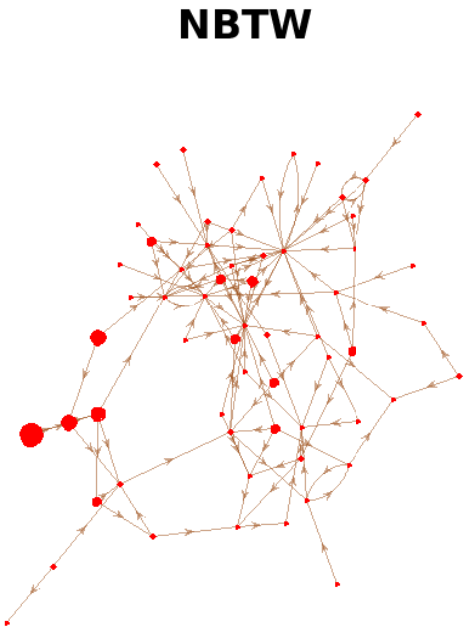


FIG. 7. Representation of the network PAJEK/GLOSSGT where node size is proportional to NBTW centrality.

Table 1. Description of the dataset: name of the digraph, number of nodes $n$ and number of edges $m$, percentage of nodes $n_{\%}$ and edges $m_{\%}$ retained after pruning, and number of iterations required.

| NAME | $n$ | $m$ | $n_{\%}$ | $m_{\%}$ | iter |
|---|---|---|---|---|---|
| PAJEK/CALIFORNIA | 9,664 | 16,150 | 2% | 3% | 7 |
| GLEICH/WB-CS-STANFORD | 9,914 | 35,555 | 57% | 80% | 15 |
| SNAP/SOC-EPINIONS | 75,888 | 508,837 | 36% | 85% | 5 |
| SNAP/WIKI-TALK | 2,394,385 | 5,021,410 | 4% | 29% | 4 |
| SNAP/CIT-PATENTS | 3,774,768 | 16,518,947 | 0% | 0% | 16 |

numerical tests we worked on larger networks to understand the extent to which the size of the problem reduces when we remove these types of nodes from real world directed graphs. In Table 1 we display the name of the network, the number of nodes and edges in the original data, and the percentage of nodes and edges that are retained in the network when sources, dangling nodes and reciprocal leaves are successively eliminated from each network. We will denote these quantities by $n_{\%}$ and $m_{\%}$, respectively. Moreoever, we display the number of iterations required to complete the pruning process.

Generally, we see that significant reductions were achieved in all cases, with the linear system dimension, $n$, reduced to at least 57% of its original value. The network SNAP/CIT-PATENTS, which is the one requiring the highest number of iterative steps, retained no nodes. All dangled or became source nodes. This can be seen from Figure 8, where we display on a semi-logarithmic scale the number of each of the three types of nodes that are removed at each iteration for all the networks in our dataset. We conclude that SNAP/CIT-PATENTS does not contain nodes connected through a reciprocal edge to any other node and, since it prunes to zero, we also conclude that, as one would expect from the nature of the dataset, none of its nodes is involved in a directed cylce of any length. (We also note that the symmetrized version of this network, where all edges are regarded as being undirected, is not a tree, so the original network is not a directed tree.)

From this set of tests we conclude that the approach of iteratively removing sources, dangling nodes, and reciprocal leaves can be of computational benefit, especially for computing or estimating the spectral radius of the matrix $C$. For the networks tested here, in the worst case the order of the matrix whose spectral radius we need to estimate is reduced from $3n$ to $0.57 \times 3n \approx 1.7n$.

### 8.3 *Symmetrized network*

We note that a spectral community detection algorithm for undirected networks based on nonbacktracking walks was introduced and studied in [20]. In their numerical tests, the authors included results on a symmetrized version of a network from [1] that is based on directed links. Now that we have a means to deal with nonbacktracking walks on directed networks it is of interest to study the effect of such a symmetrization step. We obtained the data from the NEWMAN/POLBLOGS directory of the SuiteSparse Matrix Collection. After removal of nodes with neither outgoing nor incoming links, the network contained $n = 1224$ nodes and $m = 19022$ edges. Three iterations of pruning gave $n_{\%} = 53\%$ and $m_{\%} = 83\%$. The symmetrized version contained 16715 undirected edges, that is, 33430 nonzeros in the associated adjacency matrix. In Figure 9 we scatter plot the nonbacktracking centralities for the original (vertical axis) and symmetrized (horizontal axis) networks. The left-hand picture shows the normalized NBTW centralities for $t = 0.01$, and the right-hand picture shows the corrresponding nonbacktracking eigenvector centralities. We see that symmetrizing the digraph has significantly altered the results. Indeed, in both cases the top five nodes do not overlap and the overall correlation is weak.

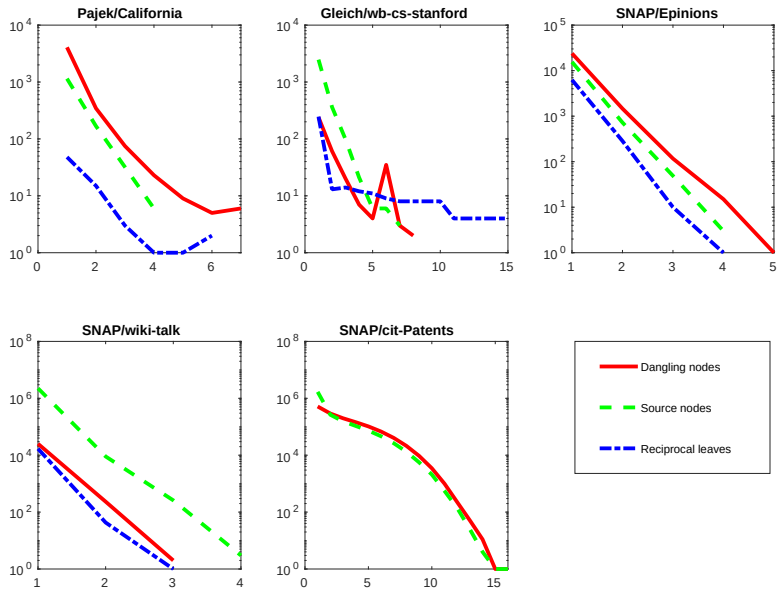


FIG. 8. Number of nodes removed at each step for each of the three types of nodes.

Overall, this test justifies the development of an algorithm that respects directed edges.

## 9. Summary

The concept of a walk around a network has proved to be extremely useful in the derivation of network centrality algorithms. In this work, we argued that some types of walk are less relevant than others. Focusing on the general case of directed networks, we showed that backtracking walks can be eliminated effectively, leading to new network centrality measures with attractive properties and, perhaps surprisingly, reduced computational cost. This approach allowed us to extend the original nonbacktracking Katz-like measure from [14], which dealt with the undirected case. We showed that the underlying generating function, and its radius of convergence, were intimately connected to a three-by-three block matrix defined in terms of the original adjacency matrix. In the undirected case, this block matrix collapses down to a two-by-two form that has appeared in previous nonbacktracking studies. From a computational perspective, a useful result is that the radius of convergence of the generating function does not change when the problem dimension is reduced by the iterative removal of certain types of nodes. Moreover, the resulting linear system for centrality may also be couched in terms of the remaining, lower dimensional, system. We showed in practice that these pruning operations can lead to significant reductions in problem size.

The treatment of nonbacktracking walk combinatorics given here has focused on network centrality issues and used the tools of matrix analysis. In future work, we plan to study further spectral properties of the generating function from a matrix polynomial viewpoint and to consider more general edge breaking and divide-and-conquer style approaches to reducing the network dimension.

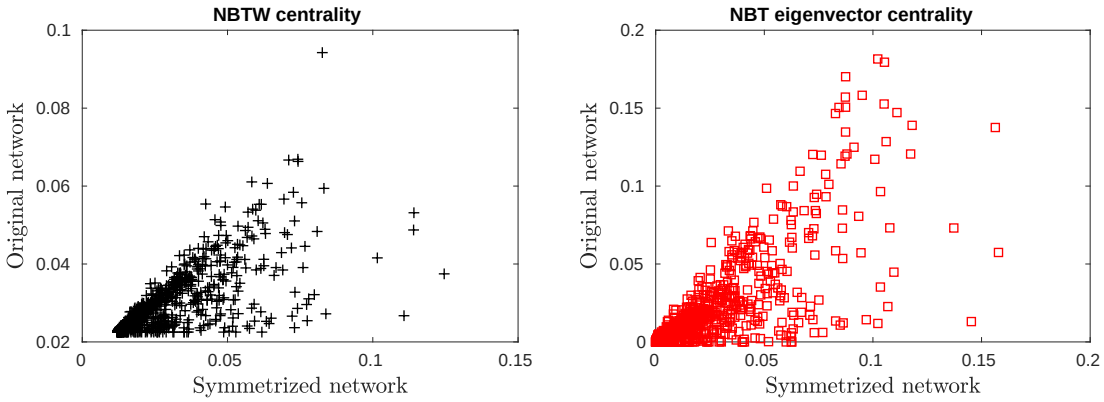


FIG. 9. Left: scatter plot of the normalized NBTW centrality computed on the original (vertical axis) and symmetrized (horizontal axis) network. Right: corresponding picture for nonbactraking eigenvector centrality.

## Acknowledgements

The work of the first and third authors was supported by grant EP/M00158X/1 from EPSRC/RCUK Digitial Economy Programme. The third author was also supported by a Royal Sociey/Wolfson Research Merit Award.

## REFERENCES

1. Adamic, L. A. & Glance, N. (2005) The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05, pages 36–43, New York, NY, USA. ACM.
2. Alon, N., Benjamini, I., Lubetzky, E. & Sodin, S. (2007) Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, **09**, 585–603.
3. Arratia, A. & Marijuán, C. (2016) On graph combinatorics to improve eigenvector-based measures of centrality in directed networks. *Linear Algebra and Its Applications*, **504**, 325–353.
4. Bonacich, P. (1972) Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, **2**, 113–120.
5. Bonacich, P. (1987) Power and centrality: a family of measures. *American Journal of Sociology*, **92**, 1170–1182.
6. Borgatti, S. P. & Everett, M. G. (2006) A graph theoretic perspective on centrality. *Social Networks*, **28**, 466–484.
7. Bowen, R. & Lanford, O. E. (1970) Zeta functions of restrictions of the shift transformation. In Chern, S.-S. & Smale, S., editors, *Global Analysis: Proceedings of the Symposium in Pure Mathematics of the Americal Mathematical Society, University of California, Berkely, 1968*, pages 43–49. American Mathematical Society.
8. Bunse-Gerstner, A., Byers, R., Mehrmann, V. & Nichols, N. K. (1991) Numerical computation of an analytic singular value decomposition of a matrix valued function. *Numer. Math.*, **60**, 1–39.
9. Davis, T. A. & Hu, Y. (2011) The University of Florida sparse matrix collection. *ACM Transactions on Mathematical Software*, **38**, 1:1–1:25.
10. Estrada, E. (2011) *The Structure of Complex Networks*. Oxford University Press, Oxford.
11. Estrada, E. & Higham, D. J. (2010) Network properties revealed through matrix functions. *SIAM Review*, **52**, 696–671.
12. Gohberg, I., Lancaster, P. & Rodman, L. (2009) *Matrix Polynomials*. SIAM, Philadelphia, PA.
13. Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J. G. & Mendes, J. F. F. (2012) Localization and spreading of

diseases in complex networks. *Phys. Rev. Lett.*, **109**, 128702.
**14.** Grindrod, P., Higham, D. J. & Noferini, V. (2017) The deformed graph Laplacian and its applications to network centrality analysis. *Preprint, submitted*.
**15.** Kato, T. (1966) *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, NY.
**16.** Katz, L. (1953) A new index derived from sociometric data analysis. *Psychometrika*, **18**, 39–43.
**17.** Kawamoto, T. (2016) Localized eigenvectors of the non-backtracking matrix. *Journal of Statistical Mechanics: Theory and Experiment*, **2016**, 023404.
**18.** Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. (2014) Multilayer networks. *Journal of Complex Networks*, **2**, 203–271.
**19.** Kleinberg, J. M. (1999) Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, **46**(5), 604–632.
**20.** Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L. & Zhang, P. (2013) Spectral redemption: clustering sparse networks. *Proceedings of the National Academy of Sciences*, **110**, 20935–20940.
**21.** Lambiotte, R., Delvenne, J. C. & Barahona, M. (2014) Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Transactions on Network Science and Engineering*, **1**, 76–90.
**22.** Langville, A. N. & Meyer, C. D. (2004) Deeper inside PageRank. *Internet Mathematics*, **1**, 335–380.
**23.** Mackey, D. S., Mackey, N., Mehl, C. & Merhmann, V. (2006) Vector spaces of linearizations for matrix polynomials. *SIAM J. Matrix Anal. Appl.*, **28**, 971–1004.
**24.** Martin, T., Zhang, X. & Newman, M. E. J. (2014) Localization and centrality in networks. *Phys. Rev. E*, **90**, 052808.
**25.** Morone, F. & Makse, H. A. (2015) Influence maximization in complex networks through optimal percolation. *Nature*, **524**, 65–68.
**26.** Morone, F., Min, B., Bo, L., Mari, R. & Makse, H. A. (2016) Collective influence algorithm to find influencers via optimal percolation in massively large social media. *Scientific Reports*, **6**, 30062.
**27.** Nakatsukasa, Y., Noferini, V. & Townsend, A. (2016) Vector spaces of linearizations of matrix polynomials: a bivariate polynomial approach. *SIAM J. Matrix Anal. Appl.*, to appear.
**28.** Newman, M. E. J. (2005) A measure of betweenness centrality based on random walks. *Social Networks*, pages 39–54.
**29.** Noferini, V. (2017) A formula for the Fréchet derivative of a generalized matrix function. *SIAM J. Matrix Anal. Appl.*, to appear.
**30.** Noferini, V. & Poloni, F. (2015) Duality of matrix pencils, Wong chains and linearizations. *Linear Algebra Appl.*, **471**, 730–767.
**31.** Pastor-Satorras, R. & Castellano, C. (2016) Distinct types of eigenvector localization in networks. *Scientific Reports*, **6**, 18847.
**32.** Stark, H. & Terras, A. (1996) Zeta functions of finite graphs and coverings. *Advances in Mathematics*, **121**(1), 124–165.
**33.** Tarfulea, A. & Perlis, R. (2009) An Ihara formula for partially directed graphs. *Linear Algebra and its Applications*, **431**, 73–85.
**34.** Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.